



# A generic model architecture for perceptual grouping

Jonas Kubilius<sup>1,2</sup>, Johan Wagemans<sup>2</sup>, & Hans P. Op de Beeck<sup>1</sup>

<sup>1</sup>Laboratory of Biological Psychology, KU Leuven (Belgium)

<sup>2</sup>Laboratory of Experimental Psychology, KU Leuven (Belgium)

get this poster & the model



klab.it

## Goal and assumptions

### Goal

Image segmentation (figure-ground assignment) using perceptual organization

### Converging lines of research

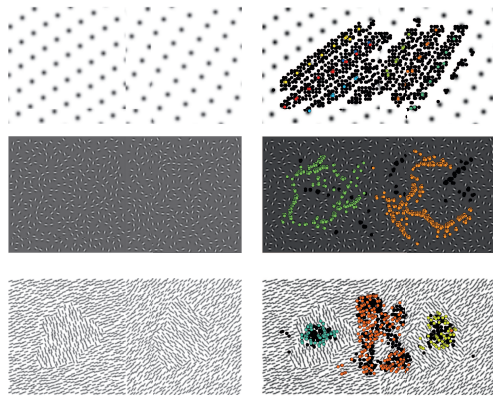
- Psychophysics and neuropsychology of spatial form defined by luminance, color, texture, motion, and binocular disparity (Regan, 2000)
- Neurophysiology on figure-ground assignment (cf. Self & Roelfsema, 2013; Kogo & van Ee, 2013)
- Perceptual organization model by Geisler & Super (2000)
- Graph-based segmentation, esp. normalized cuts (Shi & Malik, 2000)

### Assumptions

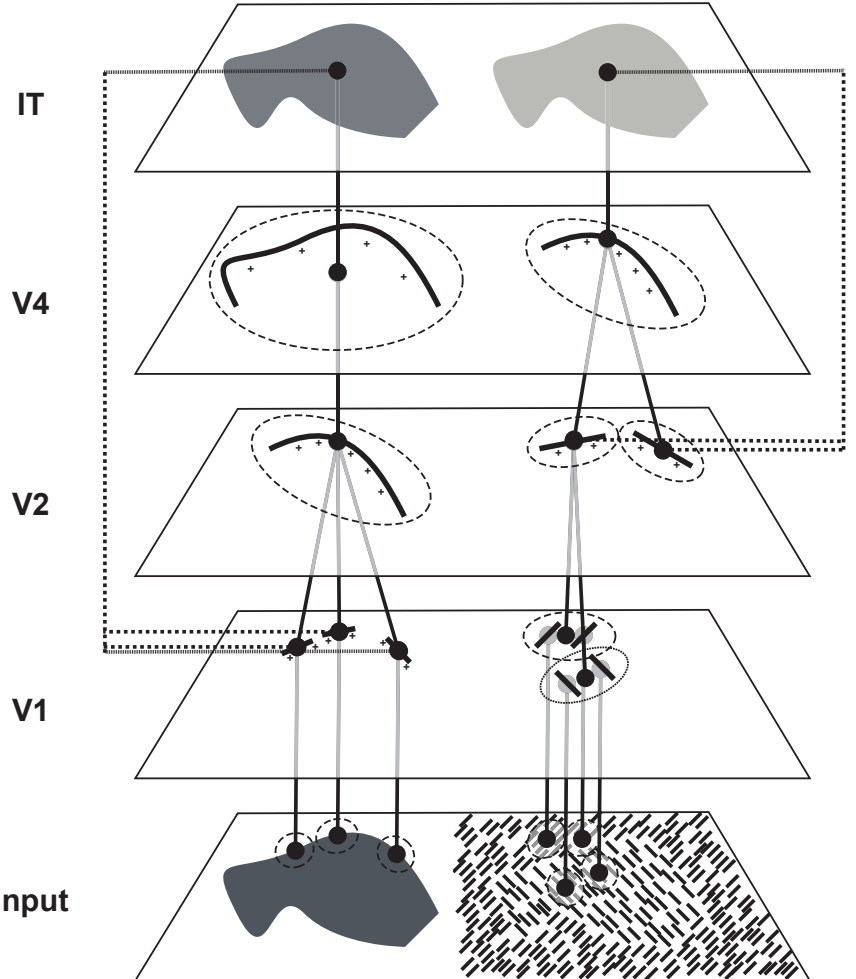
- Uniform acuity (central vision)
- Pre-attentive
- No color
- No motion
- No saccades

## Partial implementation

1. Multiscale **edge detection**
2. Initial **location sampling**: 1% maximally responding locations (3 iterations)
3. **Similarity** (grouping strength) **computation** (rules: proximity, good continuation, orientation and contrast similarity; 3 iterations)
4. Final **location sampling**: 1000 maximally responding locations
5. **Clustering** into objects using max threshold



## gmin: A hierarchy of detecting discontinuities and grouping



## Open questions

- Best way to extract features? Even edge detection is hard.
- What is the order of computations? In particular, does figure-ground assignment start before contour integration is finished?
- What is the principled way to integrate different cues?
- Excitation / inhibition: What is the grouping stopping criterion?

## Current limitations

- *Number of objects predefined*  
Equivalent to setting a grouping threshold.  
BUT: representation is hierarchical, so threshold varies depending on the task
- *Grouping principles chosen per image*  
BUT: grouping might be task-dependent (e.g., do people really see contour integration stimuli without any top-down knowledge?)
- *Location sampling density alters performance*  
Might be alleviated by Sharon et al. (2006)
- *Clustering performance not stable*  
Need a better method for clustering
- *Does not work on real images*  
Largely due to undersampling

## A note on testing models

When a segmentation model makes a mistake, is it due to:

- lack of top-down knowledge, or
  - poor bottom-up implementation?
- To avoid such confounds, top-down-free artificial scenes might be useful.



"outdoor" scene



"indoor" scene



This work was supported by a Methusalem Grant (METH/08/02) awarded to Johan Wagemans from the Flemish Government. J. K. is a research assistant of the Research Foundation - Flanders (FWO).

jonas.kubilius@ppw.kuleuven.be | klab.it | @qbilius